Personalised Travel Recommendation based on Location Co-occurrence

Maarten Clements, Pavel Serdyukov, Arjen P. de Vries, Marcel J.T. Reinders

arXiv:1106.5213v1 [cs.IR] 26 Jun 2011

Abstract

We propose a new task of recommending touristic locations based on a user's visiting history in a geographically remote region. This can be used to plan a touristic visit to a new city or country, or by travel agencies to provide personalised travel deals.

A set of geotags is used to compute a location similarity model between two different regions. The similarity between two landmarks is derived from the number of users that have visited both places, using a Gaussian density estimation of the co-occurrence space of location visits to cluster related geotags. The standard deviation of the kernel can be used as a scale parameter that determines the size of the recommended landmarks.

A personalised recommendation based on the location similarity model is evaluated on city and country scale and is able to outperform a location ranking based on popularity. Especially when a tourist filter based on visit duration is enforced, the prediction can be accurately adapted to the preference of the user. An extensive evaluation based on manual annotations shows that more strict ranking methods like cosine similarity and a proposed RankDiff algorithm provide more serendipitous recommendations and are able to link similar locations on opposite sides of the world.

Index Terms

Personalization, Location Based Services, Recommendation, Social Media, Flickr, Geotag

M. Clements is with TomTom., Netherlands. E-mail: maarten.clements@gmail.com. P. Serdyukov is with Yandex. E-mail: pavser@yandex-team.ru. A.P. de Vries is with National Research Institute for Mathematics and Computer Science (CWI). E-mail: arjen@acm.org. M.J.T. Reinders is with Delft University of Technology. E-mail: m.j.t.reinders@tudelft.nl.

I. TRAVEL RECOMMENDATION

Location based services are quickly gaining popularity due to affordable mobile devices and ubiquitous Internet access. Websites like Foursquare[1], Gowalla[2], Google Latitude[3] and Facebook[4] show that people want to share their location information and get accurate location recommendations at any time and and place. In return for sharing their location data, users can now be matched to products, venues, events or local social relations and groups.

Accurate predictions of the user's preferred locations can simultaneously aid the user itself, advertisers of products specific to the recommended place and service providers (e.g. transportation to the recommended location). To provide these recommendations, the system needs to have an accurate way to find similarities between locations or people. We propose to exploit the past visiting behaviour of people to build a location similarity model that can be used for personalised location predictions.

In this work we will exploit a set of *geotags* collected from Flickr[5] to make a recommender that can predict relevant locations for individual users. In Flickr, geotags are tuples of latitude and longitude that represent the exact location where a user made a photo. Registration of geotags can be done manually by placing the photo on a map, or automatically by the device if it is equipped with a GPS module. Here we show that the collective knowledge represented in these geotags can be used to estimate similarities between locations and that personalised location recommendations can be derived from this similarity model.

Given a user's preference in one predefined area, we predict his activity in a another disjoint area. The proposed method will be evaluated on both city and country scale and will show that places on opposite sides of the world can be related based on user location histories.

II. RELATED WORK

Since GPS equipped mobile phones have become mainstream, the amount of available geotags has grown to a number that allows for intensive data analysis. In this work, geotags are used to predict interesting locations for individual users, but the exploitation of geotags has

[1] http://foursquare.com/

[2] http://gowalla.com/

[3] http://www.google.com/latitude

[4] http://www.facebook.com/

[5] http://flickr.com/

shown to be effective for various other tasks. A method for global event detection has been proposed by Rattenbury et al., who searched for the occurrence of textual tags in spatial and temporal bursts [1]. Ahern et al. made a mapping of popular tags to geographical locations, resulting in a scale dependent map overlay with semantic information on the underlying data [2]. This work was extended by Kennedy et al. who selected relevant pictures for the predicted clusters [3]. Crandall et al. suggested not to use a fixed number of clusters and proposed a mean shift algorithm to find the most prominent landmarks and representative photos [4].

Another application of Flickr's geotags was proposed by Lee et al. who used the geographical clusters related to a tag to improve the prediction of similar tags [5]. Furthermore, several methods have been proposed to predict the geotags of a photo, based on its textual tags [6], visual information [4] and individual user travel patterns [7].

As geotags relate to a location where the user made a photo, they inherently contain a touristic preference indication. Full GPS tracks are useful to study daily mobility patterns but extra effort is needed to extract touristically interesting spots. Based on users' GPS tracks, location recommender systems have been proposed that attempt to predict popular places and activities near the current location of the user. Some work has focused on the recommendation of specific types of locations. An item-based collaborative filtering method was used to recommend shops, similar to a user's previously visited shops [8] and a user-based collaborative filtering was proposed to generate restaurant recommendations through users with similar taste [9]. Zheng et al. extensively studied GPS tracks in Beijing, defined a method to extract interesting locations from this data (*Stay regions*) and proposed a matrix factorization method to suggest locations and activities based on the current state of the user [10]. They also showed that the HITS model can effectively be used to create a ranking of popular locations and experienced people [11].

Compared to most of the previously proposed methods, our system gives recommendations in a geographically remote location, so people can use it when they are planning a trip to another country or city. We have previously showed that geotags can be used to construct a measure of similarity between locations [12]. Here, we present a thorough extension of the previous work, using a similarity model based on a scale-space of location co-occurrence data. We evaluate the potential of this similarity model for personalised recommendations. The proposed model contains a scale parameter that allows the prediction of differently sized regions. So, when a user decides to visit a certain country the recommender can be used to find the most interesting cities and when a user gets to that city the same method can be used to find the most interesting

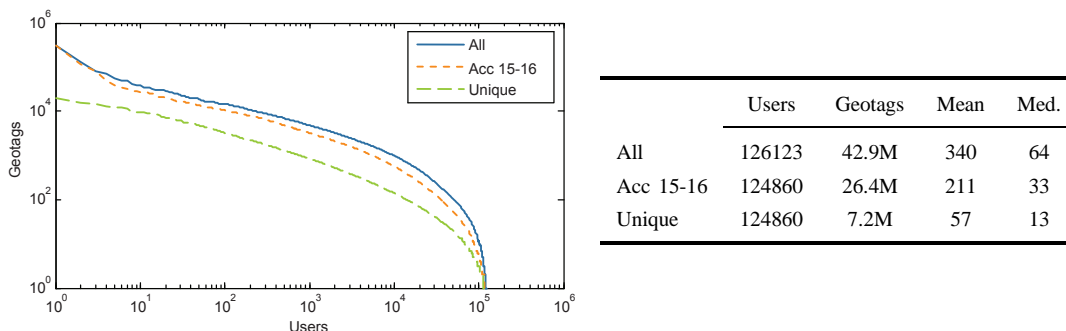|	Users	Geotags	Mean	Med.
All	126123	42.9M	340	64
Acc 15-16	124860	26.4M	211	33
Unique	124860	7.2M	57	13

Fig. 1.   The distribution of the number of geotagged photos per user in descending order. The accuracy filter reduces the data set from 43M to 26M geotags. By selecting only unique geotags we maintain 7M points. The table also indicates the mean and median number of geotags per user.

landmarks, restaurants or other venues.

Many recommendation algorithms have been proposed based on similarities between objects in a discrete item-space [13], [14], which has proven to be effective in E-commerce applications [15]. Compared to these systems, a location recommender does not have a limited number of objects to recommend. Any point consisting of two continuous values of latitude and longitude can be recommended. On a more fundamental level, we introduce a model that includes the pairwise distances between points in order to reason in this continuous space. We will demonstrate the effectiveness of this model on geographical data, but it could easily be extended to include other continuous dimensions like temporal information.

III. DATA

A. Data Collection

Using the public API of Flickr we have collected a large set of geotagged photos in a period of several months at the end of 2009 and early 2010. Figure 1 gives the distribution of the number of geotags per user (*All*). The distribution clearly shows that our crawl has a bias to people with many geotags, as the expected long tail of the distribution is missing. However, as we will only evaluate recommendations for users who have provided a sufficient amount of data, this bias in the crawl does not interfere with the objectives of this work. The total set corresponds to roughly 46% of the 93 million publicly available geotags in Flickr at the end of 2009[6].

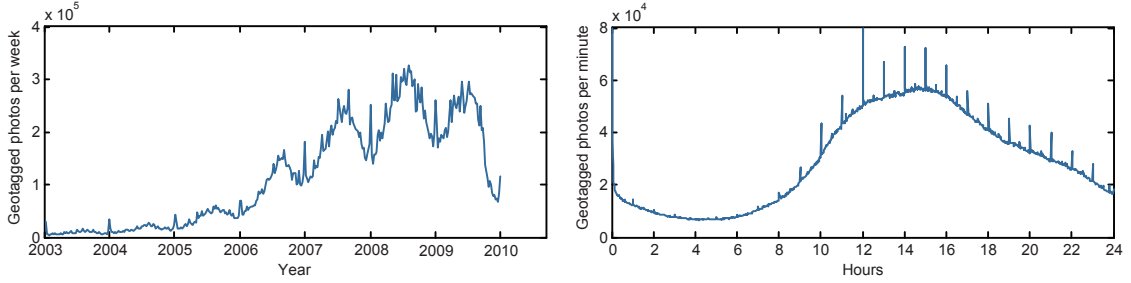[6]According to: http://www.flickr.com/map

Fig. 2.   When Flickr users make photos. Left: Photo count per week from 2003 to 2010. Right: Photo count per minute of the day, aggregated over all days.

Each geotag has an associated level of accuracy in the range of 1-16, 16 being the most accurate. This accuracy roughly relates to the zoom level of the map interface in Flickr. Because we want to make accurate predictions at the scale of individual landmarks, we keep only geotags at accuracy 15 or 16 (street level). The remaining data is represented by *Acc 15-16* in Figure 1. The possibility to integrate the accuracy value in the recommender system will be discussed in Section VIII.

Flickr allows users to upload and annotate photos in *batches*. When someone uses this function it can either mean that he made many photos at that location, or that he did not take the effort to give the exact coordinates for each individual photo. Because of the uncertainty about the user's intent when uploading a batch to a single location, we choose to ignore the possible relation between user preference and batch size and store only one geotag per batch. After these filtering steps, we retain 7.2 million geotags contributed by 125 thousand users (*Unique* in Figure 1).

B. Data statistics

The collected data set gives an interesting insight in the common behaviour of Flickr users. Besides the location of photos, Flickr also stores the date and time a photo was taken (according to the internal camera clock). Figure 2 shows the number of photos taken in a certain week between 2003 and 2010. Apart from the clear increase in popularity over the last 5 years it is interesting to see that most of the photos are taken during the northern hemisphere summer.

When we aggregate over all days and count the number of photos for each minute, we clearly see the bulk of photos is made late in the morning or early afternoon. In the evening the number of photos slowly decays until the minimum is reached around 4:30. The spikes at

full hours and at January 1st in the weekly histogram are caused by default values of empty fields in Flickr's database.

Figure 3 gives the geographical distribution of the data. This 2000x4000 histogram of the geotags clearly shows the most popular travel areas in the Flickr community. Europe and North America have the largest density of data points, but the rest of the world is also recognisable. Figure 4 gives a closer view of North America, which shows that coastlines, cities and even highways are clearly represented in the data.

Based on this data, we select the 10 most popular countries and 10 most popular cities to evaluate the feasibility of personalised travel recommendation. We rank the cities and countries by the number of users that have been there (Table I), based on their geotags located within city bounding boxes[7] and country polygons[8]. Because the number of users in the USA is much larger than other countries, we split the USA in 3 regions: East USA (Longitude $> -98.583°$), West USA (Longitude $< -98.583°$), Alaska (Latitude $> 50°$).

IV. EXPERIMENTAL SETUP

Figure 5 presents the experimental setup and the notation described in the following sections is summarised in Table II. The data is comprised of a set of users $u \in U$ who have all visited

[7]Collected in January 2010 from http://developer.yahoo.com/geo/geoplanet/

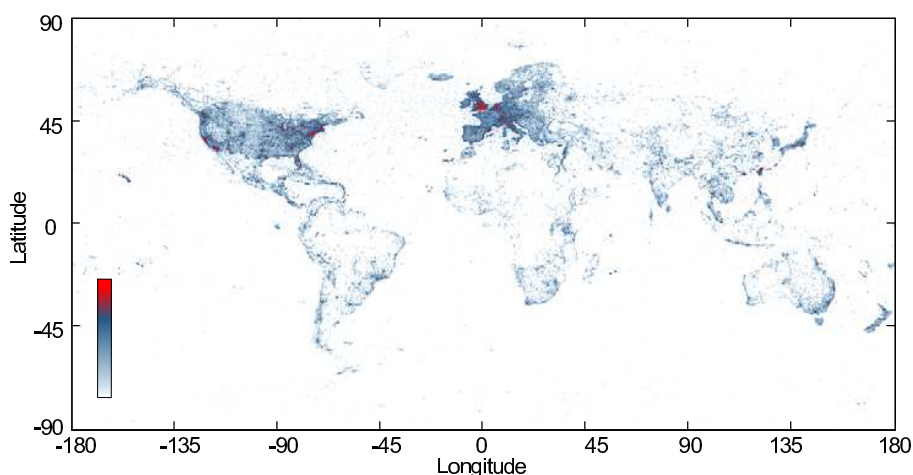[8]Collected in March 2010 from http://mappinghacks.com/

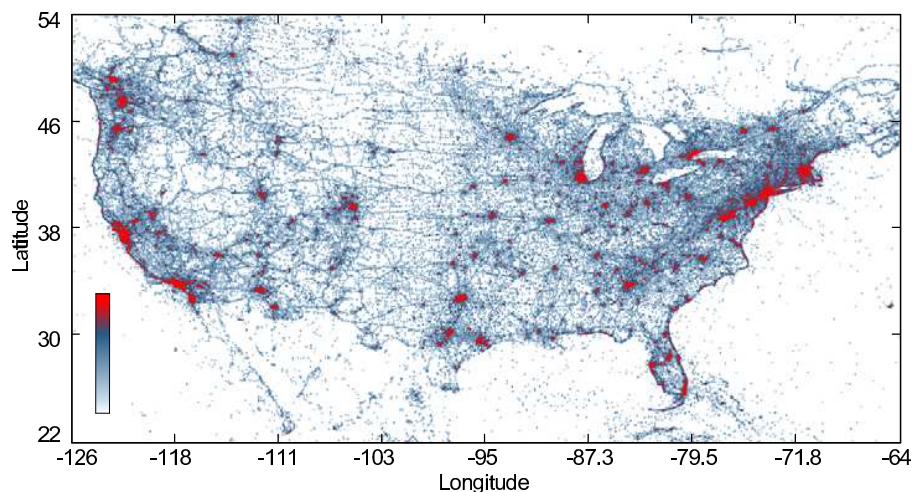Fig. 3.   Where Flickr users make photos: World distribution.

Fig. 4.   Where Flickr users make photos: USA distribution

TABLE I

NUMBER OF USERS IN TOP-10 CITIES AND COUNTRIES

Users	City	Users	Country
19802	London, England, United Kingdom	45738	United States EAST
18291	New York, NY, United States	32904	United States WEST
13786	Paris, Ile-de-France, France	25934	United Kingdom
12470	San Francisco, California, United States	18247	France
7893	Rome, Lazio, Italy	16995	Italy
7627	Los Angeles, California, United States	15414	Spain
7208	Washington, District of Columbia, United States	13381	Germany
7158	Chicago, Illinois, United States	11024	Canada
7069	Barcelona, Catalonia, Spain	6503	Netherlands
6569	Berlin, BE, Germany	5067	Australia

at least one location $l \in \mathcal{L}$, where l is a tuple (x, y, z) of Cartesian coordinates and $\mathcal{L} \subset \mathbb{R}^3$ is the set of all geotags in our data set. The set of geotags $\mathcal{L}$ is a subset of the world $\mathcal{W}$ described by a sphere with radius 6,367,449 m centered at zero. While Flickr provides the geotags in latitude and longitude we will use Cartesian coordinates throughout this work, which is more efficient for the computation of Euclidean distances between points. The distance between two points is measured through the crust of the Earth instead of over the surface. This difference is negligible for small distances and rank equal in general.
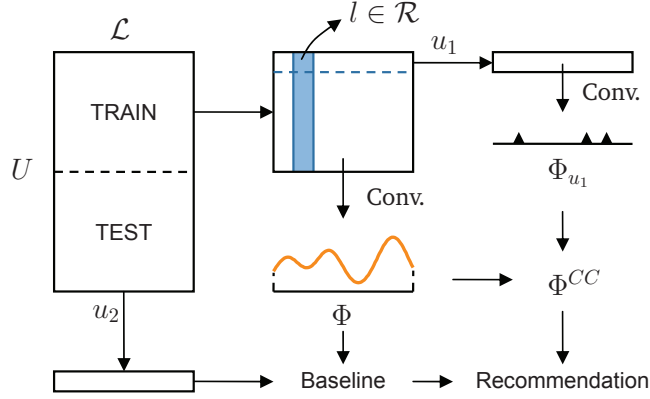
Fig. 5. Experimental setup. The training users generate the global travel distribution Φ and the location similarity model Φ^{CC}. The performance of both models for location recommendation in a predefined region $\mathcal{R}$ is evaluated on the test users.

The data from half of the users (the *training* set) will be combined in a model that captures the similarities between the most important locations in two regions. With the data from the other half of the users (the *test* set) the application of the learned co-occurrence model for personalized travel recommendations will be evaluated. We split the data in equally sized training and test sets by first ranking all users according to the number of geotags. In this order, we select users $1, 4, 5, 8, 9, \ldots$ as training users and $2, 3, 6, 7, 10, \ldots$ as test users, so the two sets will roughly follow the same distribution.

TABLE II

NOTATION USED IN THIS PAPER. FOR ALL $l, \mathcal{L}, f, \Phi, p, \mathcal{P}$ WE USE THE SUPERSCRIPT $\ldots^{s/t}$ TO REFER TO THE REGION OF THE DATA ($\mathcal{R}^s$ OR $\mathcal{R}^t$) AND THE SUBSCRIPT $\ldots_{u_k}$ IF THE DATA IS BASED ON A SINGLE USER. THE LOCATIONS IN THE CO-OCCURRENCE SPACE ($c, \mathcal{C}$) CAN ALSO CONTAIN THE SUBSCRIPT $\ldots_{u_k}$, BUT NO SUPERSCRIPT.

$u_k \in U$	The users in the Flickr data set
$\mathcal{W}$	The world; subspace of $\mathbb{R}^3$
$\mathcal{R}^s, \mathcal{R}^t$	Starting region, target region; Subspaces of $\mathcal{W}$
$l \in \mathcal{L}$	All geotags in the data set, subset of $\mathcal{W}$
f	Function describing a set of geotags
Φ	Function describing the Gaussian convolution of f
$p \in \mathcal{P}$	The peaks of Φ
$c \in \mathcal{C}$	Points in the co-occurrence space; Subset of $\mathbb{R}^6$

The objective of this work is to predict the visited locations of a test user $u_k \in U$ in a target region $\mathcal{R}^t \subset \mathcal{W}$, based on the geotags of that user in a starting region $\mathcal{R}^s \subset \mathcal{W}$. A *region* $\mathcal{R}$ can refer to either a city or a country from Table I. To evaluate the performance of the location prediction we remove all the geotags of u_k that lie within $\mathcal{R}^t$ and use the geotags of u_k in an other region $\mathcal{R}^s$ to predict the location of the removed data. For this evaluation setup we need users that have visited at least 2 distinct regions. Obviously, when the recommender is operational, recommendations can already be made when a user has visited a single region.

To build the location similarity model between $\mathcal{R}^s$ and $\mathcal{R}^t$, we first find the most popular locations in these two regions. We use a kernel convolution of the training data with a Gaussian kernel to smoothly cluster the geotags that are near to each other (Section V). We also find the most important locations per user by computing the kernel convolution over only the user's geotags. Both resulting distributions (Φ, Φ_{u_k}) are combined in the co-occurrence space Φ^{CC} which estimates the relations between the top locations in both regions (Section VI). The model Φ will be used to generate a baseline ranking (Section VII-A), the model Φ^{CC} will be used to predict a personalised location ranking per user (Section VII-B-VII-C).

V. PEAK FINDING (Φ, Φ_{u_k})

The geotags of all users are described by the function f which has a Dirac delta pulse at the locations where one of the users created a geotag and zero otherwise:

$$f(z) = \sum_{l \in \mathcal{L}} \alpha_l \delta(z - l) \tag{1}$$

where α_l is a parameter that allows the assignment of different weights per geotag. In this work α_l will be set to 1 for all l, other weighting strategies will be discussed in Section VIII.

We propose to use a Gaussian kernel convolution to obtain a smooth estimate of the density of all photos on the planet $\Phi_\sigma = f * g_\sigma$, where the Gaussian kernel is described by $g_\sigma(z) = e^{-\|z\|^2/2\sigma^2}$, for $z \in \mathbb{R}^3$. The standard deviation σ is used as a scaling parameter (or bandwidth) which gives the opportunity to set the size of the recommended locations. We do not use the common normalisation parameter of a probability density estimation with Gaussian kernels ($1/n\sqrt{2\pi\sigma^2}$, with n the number of data points) so that the convolution result will directly estimates the total number of photos taken at a certain location instead of the probability. In the rest of this work, we will drop the subscript σ for readability.

In the same way the function describing the geotag profile of a single user u_k is given by:

$$f_{u_k}(z) = \sum_{l \in \mathcal{L}_{u_k}} \alpha_l \delta(z - l) \tag{2}$$
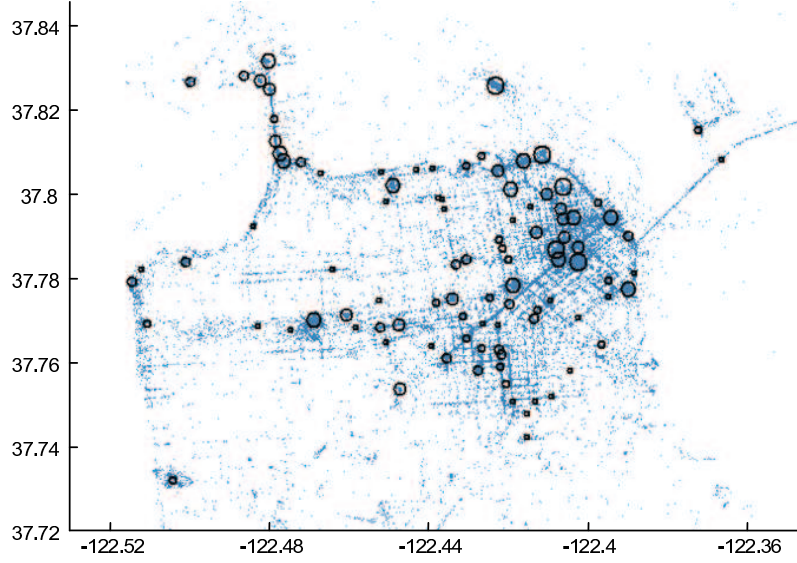
Fig. 6. The circles indicate the top-100 peaks in San-Francisco at $\sigma = 100\,\mathrm{m}$ where the radius is related to the peak amplitudes. The underlying data points clearly show the structure of the touristic part of the city.

And the density estimate $\Phi_{u_k} = f_{u_k} * g$.

We use $\mathcal{P}$ and $\mathcal{P}_{u_k}$ to denote the local maxima or *peaks* of Φ and Φ_{u_k} respectively. These peaks represent the most popular locations for all or a single user. A *mean shift* procedure is used to efficiently find the peaks of the functions [16]. We evaluate the peaks at 19 values of σ evenly distributed on a logarithmic scale from $10\,\mathrm{m}$ to $10\,\mathrm{km}$ for cities and $1\,\mathrm{km}$ to $1000\,\mathrm{km}$ on country scale. To ensure that all local maxima are found, we initiate the mean shift procedure with all individual geotags for computation on the finest scale. On each subsequent scale σ, we use the peaks from the previous scale as seeds. This procedure results in a scale-space that represents the structure of the data and allows us to analyse it at various scales.

The peaks $p \in \mathcal{P}$, found by the mean shift procedure on all geotags, can now be ranked based on their amplitude to obtain a popularity ranking of the locations in region $\mathcal{R}$ at scale σ. The application of the mean shift algorithm on geotag data was already proposed by Crandall et al. Compared to their work our scale-space will be more accurate because we use Cartesian coordinates instead of mapping latitude and longitude in a 2D plane [4]. Also, our method differs from Crandall et al. as we use a Gaussian kernel instead of a uniform disk. The Gaussian kernel convolution results in a smooth density estimate and does not generate plateau peaks. Other notable similar methods to define a popularity ranking of all locations in a given area are the
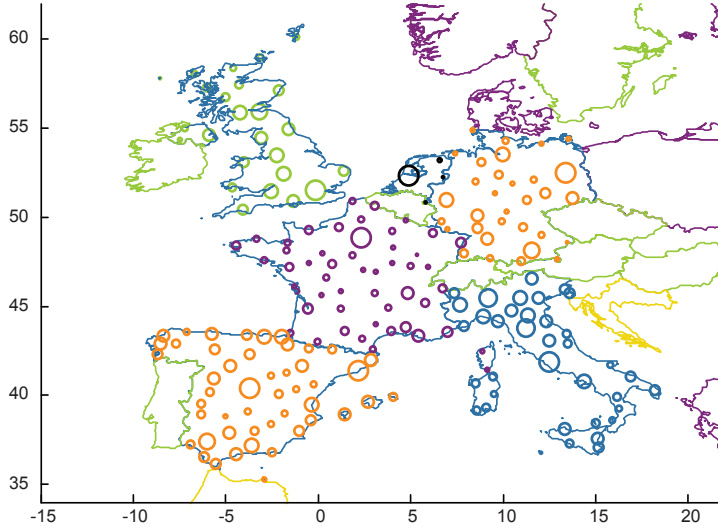
Fig. 7.　The polygons of the European countries in the top-10 most visited (Blue), top-20 (Green), top-30 (Purple) and top-40 (Yellow). The circles indicate the peaks in the top-10 most visited countries with $\sigma = 21.5\,\text{km}$, the radius is related to the peak amplitudes.

scale specific clustering in Yahoo!'s World Explorer [1], [2] and the tree-based hierarchical graph used in Microsoft's GeoLife project [11]. We chose to use the Gaussian scale space as it has a strong theoretical foundation [17] and will show to provide a logical solution to the co-occurrence model.

In Figure 6 the data points of the training users in the city center of San Francisco are shown (the actual bounding box used in this work is larger). The top-100 peaks with largest amplitude at $\sigma = 100\,\text{m}$ are depicted by circles. The clustering shows that the proposed model does capture most of the well known landmarks like *Alcatraz*, *Union Square Park*, *Coit Tower*, *Yerba Buena Gardens* and *Pier 39*. Long stretched landmarks like the *Golden Gate Bridge*, are not represented by a single cluster but several clusters appear at the popular view points. Figure 7 shows the country polygons in western Europe and for the countries in the top-10 the clusters are shown at a scale of $\sigma = 21.5\,\text{km}$. Most of the main cities are clearly visible on the map. The west of the Netherlands is grouped into a single cluster at this scale, which is reasonable as it is often seen as a single metropolitan area. At smaller scales the individual cities appear.

For computational efficiency we will only experiment with the top-500 peaks in each region. To check whether we are missing any important peaks in this step we look at the peak amplitude
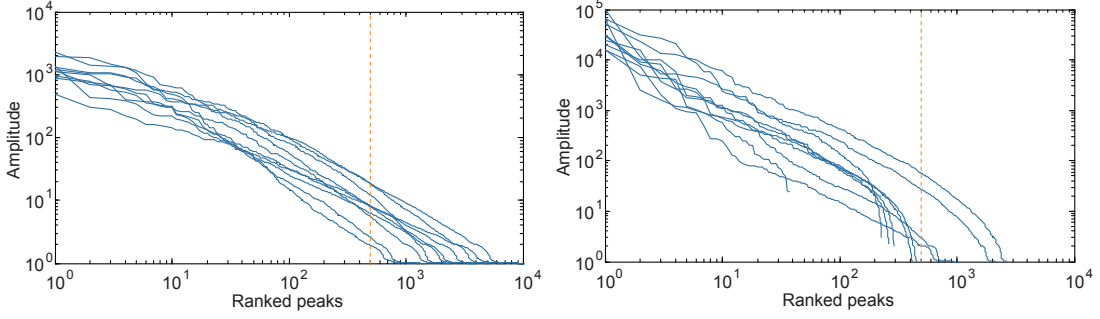
Fig. 8. The distribution of peak amplitudes at the smallest scales that will be used for evaluation in cities and countries. Left: Each line shows the peak amplitudes in one of the top-10 cities at $\sigma = 46$ m. Right: Each line represents one of the top-10 countries at $\sigma = 6.8$ km. The dotted lines indicate the cutoff at 500 peaks.

of the 500th peak in Figure 8. As the contribution of each geotag to a peak ranges between 0 and 1, the peak amplitude estimates the number of photos taken there. Because a user can make multiple photos at a single location, the number of users that contribute to the peak will be smaller: users < photos $\approx$ Peak amplitude.

The values chosen for σ will be explained in Section VII-A. At $\sigma = 46$ m there are only three cities where the 500$^{\text{th}}$ peak has an amplitude larger than 10 (London, New York, San Francisco). There are two countries (USA East and USA West) that still have large peaks after the top-500 (Amplitudes: 57 and 28). We believe that a cluster smaller than 10 photos is insignificant for our task and conclude that in most regions no important locations will be lost due to the selection of the top-500 peaks.

VI. CO-OCCURRENCE MODEL (Φ^{CC})

When visiting a country or city, most users actively plan their trip and choose the landmarks to visit based on their interests. Especially, making a photo at a certain location is a clear indication of interest in that location. Based on these assumptions, we propose to estimate the similarity between two location by the number of users that have made a photo at *both* places. As geotags are continuous points in $\mathcal{W} \subset \mathbb{R}^3$, a method needs to be found that counts the contribution of each of these points to a pair of landmarks.

We propose to create the location co-occurrence model between two regions $\mathcal{R}^s$ and $\mathcal{R}^t$ as follows. At a chosen scale σ the locations visited by u_k are selected by taking his peaks $p^s_{u_k} \in \mathcal{P}^s_{u_k}$ from $\mathcal{R}^s$ and $p^t_{u_k} \in \mathcal{P}^t_{u_k}$ from $\mathcal{R}^t$. The location co-occurrences for this user between the two regions are given by $c_{u_k} \in \mathcal{C}_{u_k}$, where $\mathcal{C}_{u_k} = \left\{ \langle p^s_{u_k}, p^t_{u_k} \rangle \,|\, p^s_{u_k} \in \mathcal{P}^s_{u_k}, p^t_{u_k} \in \mathcal{P}^t_{u_k} \right\} \subset \mathbb{R}^6$

is the set of all pairwise combinations of this user's peaks in both regions. The points in the co-occurrence space are visualised for two users by the black triangles in Figure 9.

When all the peaks of all users are added to this co-occurrence space, the most dense regions represent location pairs that are often visited by the same users, and therefore indicate a strong similarity between the two locations. A smoothed prediction of location similarities can now be derived by computing the kernel convolution over the co-occurrence space, which will be denoted as Φ^{CC}. However, since this space may contain millions of 6 dimensional data points, applying the mean shift algorithm to find the local optima is computationally expensive.

However, the locations of the most prominent landmarks are already known from $\mathcal{P}^s$ and $\mathcal{P}^t$. Therefore we only need to evaluate the value of Φ^{CC} at the pairwise location combinations from $\mathcal{P}^s$ and $\mathcal{P}^t$, visualised as orange circles in Figure 9. For example, when p_m^s and p_n^t are two peaks in Φ^s and Φ^t respectively, and the combined location is given by $c_{m,n} = \langle p_m^s, p_n^t \rangle \in \mathbb{R}^6$, the co-occurrence of these two landmarks is defined by the sum over all user contributions:

$$\Phi^{CC}(c_{m,n}) = \sum_{u_k \in U} \sum_{c_{u_k} \in \mathcal{C}_{u_k}} e^{-d(c_{m,n}, c_{u_k})/2\sigma^2} \tag{3}$$

where $d(c_{m,n}, c_{u_k})$ is the Euclidean distance between the evaluated landmark combination $c_{m,n}$ and c_{u_k} is a location co-occurrence in the profile of u_k. As we have limited the number of peaks per region to 500 there will be maximally 250,000 evaluation points per combination of $\mathcal{R}^s$ and $\mathcal{R}^t$.

The upper left point in the co-occurrence space example in Figure 9 illustrates that peak intersections from Φ^s and Φ^t may exist that do not generate a peak in the co-occurrence space Φ^{CC}: if two locations are simply never visited by a single user, the co-occurrence will be zero.

We illustrate the computation of Φ^{CC} at the bottom right evaluation point in Figure 9. Three user points contribute significantly to the co-occurrence peak, although also the small contributions from the other peaks are taken into account. The illustration also indicates that the actual peak in the co-occurrence space might be slightly shifted to a different location. The impact of the error introduced by this approximation is discussed in Appendix A.

VII. RESULTS

A. Baseline Optimisation and Evaluation Criteria

As a baseline, the peaks in $\mathcal{R}^t$ will be ranked on the score determined by the general popularity: $S(p_n^t) = \Phi(p_n^t)$. This results in a static ranking, equal for all users. After ranking the locations, we compute the distance of each of the recommended locations to the nearest
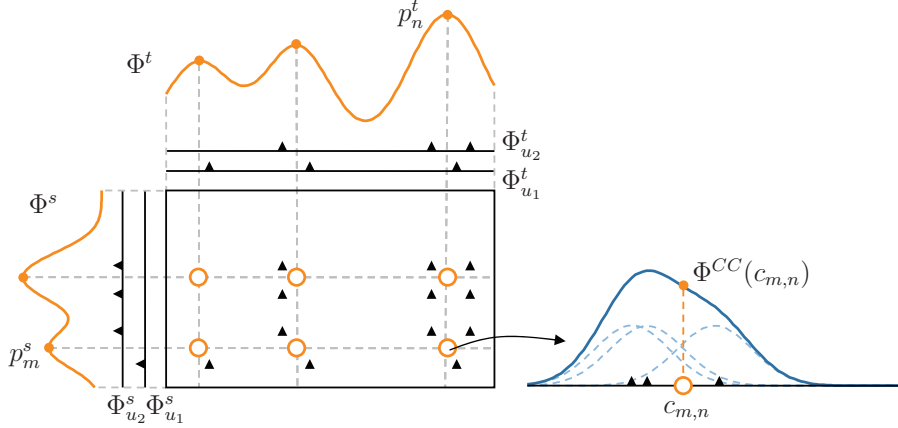
Fig. 9. Co-occurrence model. Each user's peaks are mapped into the co-occurrence space (visualised for two users). At the Top-500 peak locations of the prior distribution Φ the result of the kernel convolution in the co-occurrence space Φ^{CC} is evaluated. For one point the contribution from the two users is visualised. For visualisation purposes the 6D co-occurrence space is visualised in 2D (left) and 1D (right).

peak of the test user in $\mathcal{P}_{u_k}$ (at the same σ). We then set a threshold PC on this distance and count a recommended location as correct if the nearest of the user's peaks lies within this threshold. At small scale values, many peaks will be predicted close to each other. To make sure the recommender does not get rewarded for the suggestion of a single landmark multiple times, we disqualify a recommended location if it lies within distance PC from an earlier prediction.

The predicted location ranking will be evaluated on four criteria:

Precision (P@5), defined as the fraction of correct recommendations in the top-5.

Mean average precision (MAP@50), the mean over the precision values after each correct recommendation in the top-50.

NDCG$_{\text{IP}}$. Similar to Zhou et al. we want to express the *surprisal value* of the recommended list in a number [18]. We propose to use the Normalised Discounted Cumulative Gain (NDCG) by Järvelin and Kekäläinen which compares the predicted ranking to the optimal possible ranking [19]. The NDCG allows the assignment of a *gain* value to account for differences in relevance between the ranked objects (please refer to [19] for details). To measure the surprisal value of the predicted ranking we set the gain of each correctly recommended location p_n^t to the inverse popularity $1/\Phi(p_n^t)$ abbreviated as IP, so that less popular locations contribute more to the result than
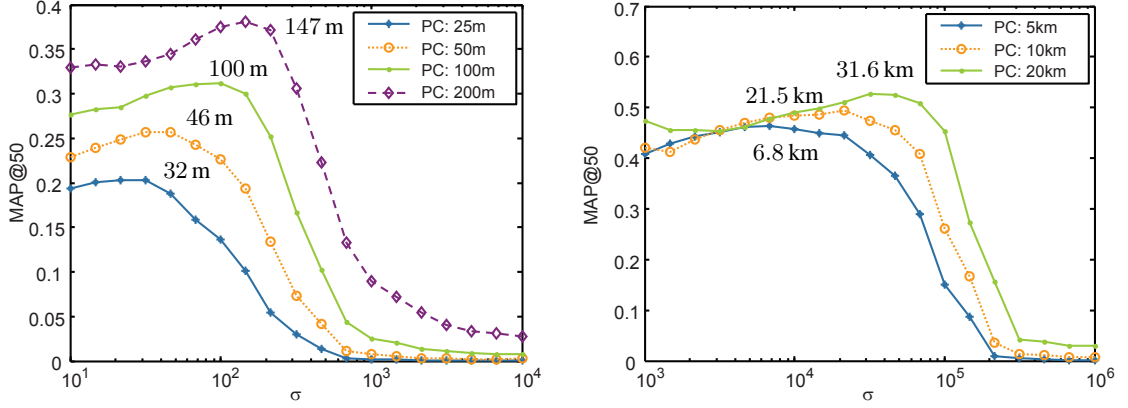
Fig. 10. Performance of the baseline ranking using MAP@50. Left: Results on city scale, for the full range of σ and $PC \in \{25, 50, 100, 200\}$ m. Right: Results at country scale for $PC \in \{5, 10, 20\}$ km.

popular locations. Then we compute NDCG$_{\text{IP}}$ over the resulting ranking. The optimal NDCG$_{\text{IP}}$ will be obtained when we correctly predict all the user's test locations, but in reverse order of popularity.

Benefit ratio (BR), the number of users who get an improved recommendation over the baseline divided by the number of users who get a deteriorated recommendation. BR can be computed over any of the previously defined evaluation methods.

To only evaluate users who have provided a decent amount of preference information, we consider those users who have at least 5 peaks at the lowest level of the scale-space ($|\mathcal{P}_{u_k}| \geq 5$). At city scale this pruning step means that users must have at least 5 peaks in Φ_{u_k} at $\sigma = 10$ m. At country scale, users need to have at least 5 peaks in Φ_{u_k} at $\sigma = 1$ km.

The optimal σ at a chosen value of PC will be estimated based on MAP@50. Compared to P@5, the results on MAP@50 more gradually change with different values of σ, therefore parameter optimisation on MAP@50 gives a more reliable estimate of the optimal setting. P@5 however gives a more intuitive evaluation on the practical usability of the recommender. We will therefore show the results on both criteria in the next sections.

In Figure 10 the mean MAP@50 is plotted for the baseline ranking for the full range of σ values and various settings of PC. For all settings, the choice of σ has a clear optimum. When σ is chosen too small, multiple peaks exist at a single landmark, while for too large σ individual landmarks will be missed because they are merged into a single peak. At city scale the optimal σ is found close to the selected value of PC. At country scale we find that the

optimal σ is larger. This can be explained by the fact that within a city the ratio between the point of interest size and the distance between them is larger than in a country.

At both city and country level, we select two scales for further evaluation. Within city recommendation will be evaluated at $PC = 50\,\mathrm{m}$, $\sigma = 46\,\mathrm{m}$ and $PC = 100\,\mathrm{m}$, $\sigma = 100\,\mathrm{m}$. At country scale we will evaluate recommendations at $PC = 5\,\mathrm{km}$, $\sigma = 6.8\,\mathrm{km}$ and $PC = 10\,\mathrm{km}$, $\sigma = 21.5\,\mathrm{km}$.

B. Recommendation

1) Generating Recommendations: We compute $\Phi^{CC}(\langle p_m^s, p_n^t \rangle)$ for all paired peaks in the top-500 $p_m^s \in \mathcal{P}^s$ and the top-500 $p_n^t \in \mathcal{P}^t$ in all combinations of $\mathcal{R}^s$ and $\mathcal{R}^t$ (the top-10 cities and countries), based on the set of training users. The derived models can now be used to generate recommendations for the test users.

As explained in Section IV the geotags of test user u_k in a starting region $\mathcal{R}^s$ will be used to predict the visited locations in $\mathcal{R}^t$. The predicted location ranking in $\mathcal{R}^t$ will then be compared to the locations actually visited by u_k. In order to evaluate the performance of the predicted recommendations for a test user, the user therefore needs to have visited at least two distinct regions. In both regions we enforce the pruning settings at $|\mathcal{P}_{u_k}^s| \geq 5 \wedge |\mathcal{P}_{u_k}^t| \geq 5$ as explained in Section VII-A.

The score of location p_n^t in $\mathcal{R}^t$ for user u_k is now derived by:

$$S^{CC}(p_n^t, u_k) = \sum_{p_m^s \in \mathcal{P}^s} \sum_{p_{u_k}^s \in \mathcal{P}_{u_k}^s} \Phi^{CC}(\langle p_m^s, p_n^t \rangle) e^{-d(p_m^s, p_{u_k}^s)/2\sigma^2} \tag{4}$$

which counts the contribution of each of the user's peaks $p_{u_k}^s$ in $\mathcal{R}^s$ to each of the landmarks p_m^s in $\mathcal{R}^s$, and weights each of these landmarks with the co-occurrence model. To predict the recommendations for u_k when traveling to $\mathcal{R}^t$, the locations p_n^t are ranked according to this score and the top ranked locations are recommended. This computation is visualised for a user u_k with three geotags in $\mathcal{R}^s$ in Figure 11.

2) Recommendation performance: We now compare the ranking on S to the ranking predicted by S^{CC}. Table III contains the results at the two selected scales for between-city and between-country recommendation. The presented values are averaged over all possible recommendations for all city/country pairs in the top-10 lists. At city scale the results are based on 16,620 measurements, with an average user size of 9 locations (median). At country scale we can evaluate 13,476 recommendations, with a median user size of 7.
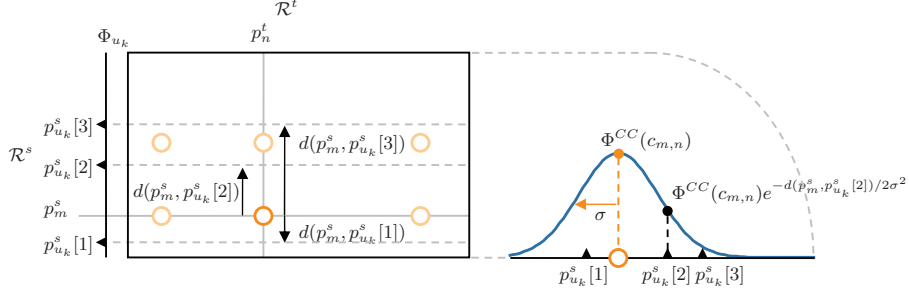
Fig. 11. Computing recommendations with the location co-occurrence model. For each peak p_m^s in $\mathcal{R}^s$ all contributions of all the user's geotags are aggregated using a Gaussian distribution as weight function. Then the final score of a location p_n^t in $\mathcal{R}^t$ is derived from the sum over all p_m^s.

For all settings and all evaluation methods our model improves over the baseline. We test the significance of the improvement using a Wilcoxon signed rank test, which tests the hypothesis that the difference between the matched samples in the two sets comes from a distribution with zero median. At a confidence level of 1% only the results on P@5 for $\sigma = 46$ m are not significant. Probably too many landmarks will be represented by multiple peaks at this scale, making the co-occurrence model less accurate.

The improved results on NDCG$_{\text{IP}}$ indicate that not only the rank position of the test results improves, but also the surprisal value of the presented recommendations. The co-occurrence

TABLE III

RESULTS OF THE BASELINE (S) COMPARED TO THE RECOMMENDER (S^{CC}), FOR TWO SCALES AT BOTH CITY AND COUNTRY LEVEL.

	City				Country			
	$PC = 50$ m		$PC = 100$ m		$PC = 5$ km		$PC = 10$ km	
	$\sigma = 46$ m		$\sigma = 100$ m		$\sigma = 6.8$ km		$\sigma = 21.5$ km	
	S	S^{CC}	S	S^{CC}	S	S^{CC}	S	S^{CC}
P@5	0.237	0.237	0.293	0.300	0.266	0.274	0.257	0.261
MAP@50	0.311	0.312	0.370	0.377	0.437	0.445	0.482	0.488
NDCG$_{\text{IP}}$	0.237	0.238	0.272	0.277	0.287	0.293	0.358	0.365
BR-P@5	1.034		1.375		1.419		1.337	
BR-MAP@50	1.046		1.246		1.248		1.298	
BR-NDCG$_{\text{IP}}$	1.108		1.361		1.292		1.476	

model gives better performance while less popular locations are observed at the top of the ranking. This shows that the method correctly learns how the preference of the user differs from the average.

Although BR shows a decent improvement when the recommendation model is used, the mean absolute improvement on the individual evaluation criteria is small. For many users the popularity based baseline and the personalised ranking of recommended locations are very similar. Two reasons can be given for these small differences. First, many users do not have a single preference (e.g. only visit botanical gardens), but visit many types of landmarks when they come to a new location. With the proposed co-occurrence model, the combined recommendations based on these mixed preference profiles converge to the prior ranking. Second, because many people visit the most popular locations in the target region the evaluation method actually expects us to recommend these. This is inherent to the evaluation of recommendations with a train and test set.

In Section VII-C we will see that when a single type of landmark is used as starting location and we manually asses the recommended locations, the prediction is highly accurate and we can use more extreme weighting methods to exploit the location co-occurrence.

3) Tourist Filter: We hypothesize that people who visit both $\mathcal{R}^s$ and $\mathcal{R}^t$ for touristic purposes will benefit more from the recommendations than people who actually live in one of the cities. To confirm this hypothesis we implement a tourist filter as follows: Based on the creation date of the photos in the Flickr data a user qualifies as tourist in a certain city if all his photos in that city are taken in n periods of 14 days. So in the 3x14 filter we allow the user to visit a single city 3 times, and all the user's photos have to be taken in at most 3 different windows of 14 days.

The results with three different tourist filters applied in both $\mathcal{R}^s$ and $\mathcal{R}^t$ are presented for $\sigma = 100\,\mathrm{m}$ in Table IV. First, we observe that both the baseline and the recommendation performance go up when a more stringent filter is used. So tourists conform more to the overall visiting behaviour than city inhabitants. Second, when we set a more strict tourist filter, the performance difference between the recommender and the baseline goes up. This means that touristic behaviour in one city should be predicted by touristic behaviour in another city.

Table IV also indicates the number of recommendations (*Recs*) that can be evaluated with each filter. We need a user two have made a touristic visit in at least two different cities in order to evaluate the performance. These two criteria cause the number of evaluations to drop quite quickly.

TABLE IV

RESULTS OF CITY SCALE RECOMMENDATION AT $\sigma = 100$ m FOR DIFFERENT TOURIST FILTERS. THE BEST
RESULTS ARE OBTAINED FOR THE MOST STRICT FILTER (1X14).

	City, $PC = 100$ m, $\sigma = 100$ m							
	All		3x14		2x14		1x14	
	S	S^{CC}	S	S^{CC}	S	S^{CC}	S	S^{CC}
P@5	0.293	0.300	0.321	0.330	0.331	0.339	0.339	0.351
MAP@50	0.370	0.377	0.409	0.417	0.419	0.427	0.430	0.440
NDCG$_{\text{IP}}$	0.272	0.277	0.301	0.308	0.307	0.314	0.318	0.325
BR-P@5	1.375		1.511		1.491		1.687	
BR-MAP@50	1.246		1.338		1.358		1.422	
BR-NDCG$_{\text{IP}}$	1.361		1.491		1.547		1.614	
Recs	16,620		8576		6600		3536	

4) Within-City Recommendation: Song et al. showed that the every day mobility patterns of people are highly predictable 93% of the time [20]. Other related work on location prediction also focused on making recommendations close to the current location of a user [8]–[10]. We suspect that prediction of touristic behaviour in previously unvisited areas is a much harder task. First, touristic behaviour is less predictable than every day life behaviour. Second, remote predictions allow many more possibilities than nearby recommendations.

To test whether we can use our model for within-city recommendations we compute the

TABLE V

RESULTS ON RECOMMENDATION OF THE LOCATIONS FOR THE LAST DAY OF A CITY VISIT.

	City, $\sigma = 100$ m, $PC = 100$ m			
	No pruning		$|\mathcal{P}_{u_k}^t| \geq 5$	
	S	S^{CC}	S	S^{CC}
P@5	0.042	0.047	0.108	0.129
MAP@50	0.099	0.119	0.197	0.244
NDCG$_{\text{IP}}$	0.126	0.141	0.208	0.234
BR-P@5	2.452		5.182	
BR-MAP@50	1.966		2.690	
BR-NDCG$_{\text{IP}}$	1.982		2.531	
Recs	18,344		896	

co-occurrence space Φ^{CC} within each city ($\mathcal{R}^t = \mathcal{R}^s$) and set the self co-occurrence of each location to 0. For each user u_k in the test set that has been to $\mathcal{R}^t$, we cut off the last day of photos made in that city. We use the geotags created by u_k on all previous days as starting points and try to predict the user's behaviour on the final day of his stay. To split the user's data in days we use the creation date and time of the photos shifted backwards by 4.5 hours based on the results in Figure 2.

Table V gives the results at $\sigma = 100\,\text{m}$ averaged over all users (*No pruning*), and limited to users who have at least 5 peaks in $\mathcal{P}^t_{u_k}$ at this scale in both the test day and the training days. The absolute evaluation scores are lower than the scores reported in between-city recommendation, because we have fewer evaluation points in this setup. After pruning, the median user has 6 points on the test day, compared to a median of 9 in city to city recommendation.

The relative improvement with the personalised model is much larger for within-city recommendation than that for between-city recommendation. Especially for users with many geotags on the training and test day the personalised prediction clearly outperforms the baseline. Unfortunately, only few users (*Recs*) have provided enough data to pass the pruning settings. These findings indicate that adapting the location prediction to a user's personal interest is easier if the user stays within the same city.

We assume that the reason for this improvement is that users have a bias to make many photos within a certain area (e.g. close to the hotel). To verify this, we plot the probability density function (PDF) of the distance between two randomly selected geotags and the PDF of the distance between a geotag selected from the training days and a geotag selected from the test day of a single user (Figure 12). The dotted lines indicate the median of both distributions. Clearly the geotags selected from a single user have a larger probability to be close together. This location prior explains why recommendations within a single region are easier to predict than between two remote locations, confirming the second intuition given above, that remote locations allow more possibilities than nearby ones.

5) Conclusions: Because many users visit the same popular locations, prediction according to the prior travel probability is hard to improve upon. Although the absolute improvement is small, the co-occurrence model can give improved recommendations for most users.

Tourists can be selected by setting a maximum value on the number of days spent on a certain location. We find that tourists comply more with the general travel preference and are therefore more easy to predict by the baseline. Also, the relative improvement of the personalised model over the baseline is larger than for the average user, which shows that
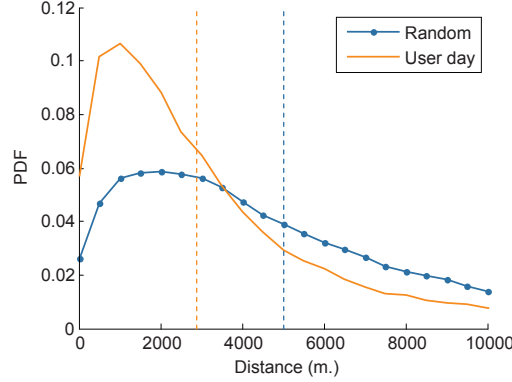
Fig. 12. PDF of distance between two *random* geotags and between the last and previous days of a single user (*user day*).

tourists have a clear preference that relates their behaviour in different cities. This shows that the location co-occurrence model based on the travel history of tourists can effectively be used to predict personalised travel recommendations. We have used a simple tourist filter and suggest that more elaborate methods could be used based on the users' profile information or textual tags.

Within-city recommendations are easier because the training data contains a location prior. If we know where the user was in the past few days, he is more likely to be in the same place the next day.

C. Serendipitous Ranking

1) Ranking Criteria: Using part of the users' real data points as test set, we have evaluated whether we can predict where the user will go if he is not influenced by a recommender. This evaluation is however strongly biased by the most popular locations in the target area. As most people will visit the *Eiffel Tower* when they get to Paris, it pays off to predict this with the recommender. However, the user would benefit more from a recommendation of a location that is not obvious and perhaps even unknown to the user. Related work on recommender systems has therefore argued that manual judgement of the recommended items is necessary for the evaluation of novel recommendations [21].

To test whether the proposed co-occurrence model can be used to produce serendipitous recommendations, we have manually annotated various sets of landmarks at city and country scale. We first use one of the landmarks (p_m^s) in $\mathcal{R}^s$ as starting point and try to predict the

annotated landmarks (p_n^t) that fall in the same category in $\mathcal{R}^t$, using the following known ranking criteria:

Prior (S) Ranking based on $\Phi(p_n^t)$.

Direct (S^{CC}) As the user profile now consists of only a single peak from Φ in $\mathcal{R}^s$, Equation 4 reduces to a ranking based directly on $\Phi^{CC}(c_{m,n})$.

Cosine (CS) Ranking based on $\Phi^{CC}(c_{m,n})/\sqrt{\Phi(p_m^s)\Phi(p_n^t)}$. Cosine similarity corrects for the popularity bias by dividing the co-occurrence by the popularity of both individual landmarks.

We also propose a new ranking method, which assigns the prior amplitudes (Φ) as weight to all locations and then compares the weight difference between the initial and new ranking:

RankDiff (RD) Let R_1 be the rank index (position in the ranked list) of a location based on $\Phi(p_n^t)$ and R_2 the rank index of the same location in $\Phi^{CC}(c_{m,n})$. Let Ψ be the list of peak amplitudes of Φ ranked in descending order. RankDiff is now defined as $RD(p_n^t) = \Psi(R_2) - \Psi(R_1)$.

The rationale behind this method is that a location that used to be at rank R_1 and had an amplitude of $\Phi(p_n^t)$ managed to reach a new ranking of R_2 where a location with amplitude $\Phi(p_x^t)$ used to be. The difference between these two amplitudes can now be seen as the amount of evidence needed to accomplish this rank gain.

Note that we also considered other ranking algorithms, that performed worse or very similar to any of the above (i.e. *Jaccard coefficient*, *Pointwise Mutual Information (PMI)*, *Lift* [22], [23]); the results of these ranking criteria are therefore left out of the discussion.

2) City scale: We manually annotate a set of baseball stadiums (Table VI) and a set of modern or contemporary art venues (Table VII) in the top-10 cities. We now repeatedly select one of the cities as $\mathcal{R}^t$ and rank all landmarks in that region based on one landmark in $\mathcal{R}^s$. As evaluation we compute the number of times a target location (from one of the two sets) goes *up* or *down* in the ranking compared to a ranking based on S, the precision at 5 (P@5), recall at 5 (R@5), defined as the fraction of correct results ranked in the top-5 and precision at R (P@R), where R is the total number of correct locations that can be recommended. For all evaluations the peaks in Φ at $\sigma = 100\,\mathrm{m}$ are used, since at this scale it is easy to manually relate each peak to a single landmark.

The results in Table VIII show that almost all baseball stadiums are related to each other as 45 out of 48 times a stadium gets a higher ranking based on co-occurrence than on the

TABLE VI

BASEBALL STADIUM SET. THE PRIOR RANK IS THE RANK INDEX BASED ON S IN THE CORRESPONDING REGION.

Stadium	City	Prior Rank	Longitude	Latitude
Yankee Stadium	New York	27	40.8271	-73.9281
City Field	New York	44	40.7557	-73.8481
Richmond Co. Bank Ballpark	New York	151	40.6457	-74.0761
AT&T Park	San Francisco	13	37.7785	-122.3896
Dodger Stadium	Los Angeles	12	34.0735	-118.2400
Nationals Park	Washington	22	38.8729	-77.0076
Wrigley Field	Chicago	5	41.9479	-87.6558
Cellular Field	Chicago	18	41.8300	-87.6340

TABLE VII

MODERN ART MUSEUM SET. THE PRIOR RANK IS THE RANK INDEX BASED ON S IN THE CORRESPONDING REGION.

Museum	City	Prior Rank	Longitude	Latitude
Tate Modern	London	4	51.5081	-0.0990
Museum of Modern Art	New York	5	40.7610	-73.9771
Guggenheim Museum	New York	12	40.7831	-73.9591
Centre Pompidou	Paris	6	48.8604	2.3520
Hirshhorn Museum	Washington	10	38.8888	-77.0230
MACBA	Barcelona	7	41.3832	2.1668
Fundacio Miro	Barcelona	28	41.3686	2.1597
Neue Nationalgalerie	Berlin	23	52.5070	13.3681
Haus der Kulturen der Welt	Berlin	21	52.5187	13.3648
Hamburger Bahnhof Museum	Berlin	28	52.5283	13.3719

prior (48 is the total number of possible ways to select two landmarks from different cities). A ranking directly based on S^{CC} does get the target locations higher in the list, but the more popular locations are often still at the very top of the ranking, resulting in a limited P@5, R@5 and P@R. The other methods make more mistakes on up/down, but RankDiff clearly improves precision and recall. The P@R of 0.46 indicates that RankDiff gets the target stadium(s) to the very top of the ranking in about half of the cases, which is a remarkable achievement given the relatively low prior rank of the stadiums.

Further inspection of the ranking produced by cosine similarity shows that many very small
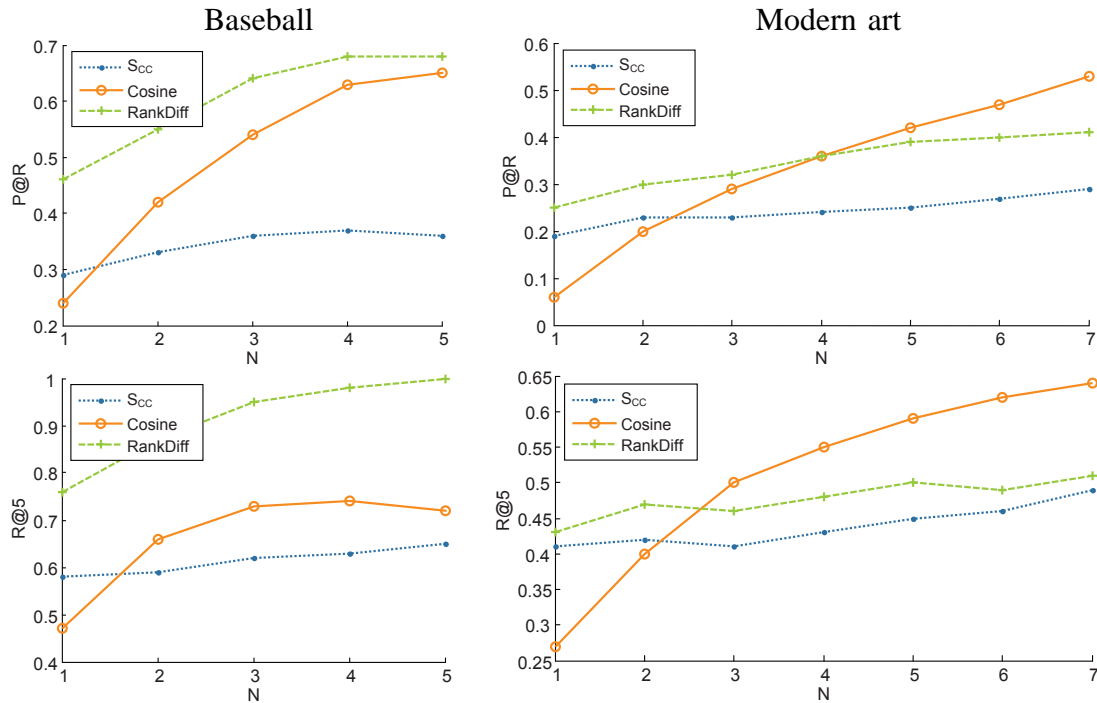
Fig. 13.   Performance difference of S^{CC}, Cosine and RankDiff with increasing profile information (N). The two left panels show the P@R and R@5 for Baseball recommendation, the right panels for Modern art.

peaks are ranked at the top. Cosine similarity can easily generate a high score when an unfamiliar starting location is used, if by coincidence the users who have been there have also another location in common. RankDiff is somewhat more conservative as it is more dependent on the absolute value of Φ^{CC} than the relative difference between Φ^{CC} and Φ.

TABLE VIII

RESULTS ON BASEBALL STADIUM AND MODERN ART PREDICTION. MARK THAT THE NUMBER OF TEST LOCATIONS IN ALL CITIES IS SMALL, THEREFORE THE MAXIMUM POSSIBLE P@5 EQUALS 0.30 FOR BASEBALL STADIUMS AND 0.32 FOR MODERN ART MUSEUMS.

Method	Baseball					Modern Art				
	Up	Down	P@5	R@5	P@R	Up	Down	P@5	R@5	P@R
S	0	0	0.04	0.09	0	0	0	0.07	0.26	0
S^{CC}	45	3	0.16	0.58	0.29	53	18	0.10	0.41	0.19
CS	41	7	0.15	0.47	0.24	30	49	0.07	0.27	0.06
RD	44	4	0.23	0.76	0.46	39	38	0.12	0.43	0.25

Although the modern art data set appears to be less coherent, the order of the methods is similar. Because many of the venues already have a high prior ranking it is hard to improve the prediction. RankDiff again gives the best performance on precision and recall.

To study the benefit of having more profile information from a user, Figure 13 shows P@R and R@5 for the three personalised methods on both data sets while the number of starting locations is increased. When two starting locations are located in different cities we simply sum the Φ^{CC} values before computing the ranking criteria. The results are averaged over all target cities and all possible combinations of N landmarks selected from the other cities.

When the recommendations from more starting points are aggregated the prediction generally gets better. The prediction of baseball stadiums based on RankDiff even reaches a R@5 of 1, meaning that in all cases the target locations are ranked in the top-5. If more information is present, Cosine similarity is less prone to mistakes and shows a steep upward trend in performance.

3) Country scale: To evaluate the co-occurrence model at country scale, we manually annotate a large set of the peaks in USA West at $\sigma = 21.5\,\mathrm{km}$ and select various starting locations in other countries to see how they influence the ranking in USA West.

Based on the prior ranking (not shown) the top-10 of locations in USA West contain 9 cities and only 1 national park (*Yosemite NP*). If we use *Ayers Rock* in Australia as starting point we expect recommendations that refer more to natural locations and less to cities. A ranking directly based on S^{CC} does show that some natural parks increase their ranking, but the co-occurrence with the top-4 cities is still larger, simply because their prior visit probability is larger (see Table IX).

We find that especially cosine similarity returns very interesting recommendations. Figure 14 and Table IX show that almost all places in the top-10 refer to rock formations in the USA, which is quite amazing since absolutely no semantic information (like textual tags) is used in the prediction.

In this example, cosine similarity seems to give better results than RankDiff. On this scale there are hardly any obscure peaks, therefore we can take the risk of using a method that can get small peaks very high in the ranking, and cosine similarity is able to get peaks from the lower part of the ranking to the top. This introduces more risk in the recommender, but can also give more interesting and serendipitous recommendations.

4) Conclusions: When the co-occurrence model is used to generate a location ranking based on a single preference point, we observe great performance increase over the prior ranking. A

TABLE IX

TOP-10 RECOMMENDATIONS BASED ON AYERS ROCK, AUSTRALIA. R IS THE NEW RANKING, PR IS THE PRIOR
RANKING (BASED ON Φ).

	S^{CC}		Cosine		Rankdiff	
R	PR	Location	PR	Location	PR	Location
1	1	San Fransisco	129	Painted Hills SP	4	Las Vegas
2	4	Las Vegas	122	Craters of the Moon NM	32	Bryce Canyon NP
3	3	Los Angeles	44	Monument Valley SP	44	Monument Valley SP
4	2	Seattle	99	Idaho Falls	36	Mt. Rushmore NM
5	32	Bryce Canyon NP	32	Bryce Canyon NP	13	Lake Tahoe
6	44	Monument Valley SP	36	Mt. Rushmore NM	14	Grand Canyon NP
7	5	Portland	62	Mt. Shasta	17	Maui
8	36	Mt. Rushmore	49	Crater lake	49	Crater lake NP
9	13	Lake Tahoe	141	Roswell	62	Mt. Shasta
10	14	Grand Canyon NP	153	Socorro / Box Canyon	122	Craters of the Moon NM

ranking based on S^{CC} directly does get the correct locations higher in the list, but not to the very top of the ranking. We find that more extreme weighting methods can be used to fully exploit the co-occurrence model.

Cosine similarity can give very small peaks as recommendations when the co-occurrence happens to be relatively large compared to the prior visiting probability. The Ayers rock example showed that this can give very interesting results. Using solely the location history of Flickr users, we were able to relate rock formations on completely opposite sides of the world.

When limited information is available the risk of recommending something unknown is high when cosine similarity is used. The proposed method *RankDiff* is more conservative, the results are more reliable but may be less surprising. On a manually annotated set of baseball stadiums we showed that the RankDiff method is able to perfectly predict where a stadium in an unvisited city is located if several other stadiums are used as starting points.

VIII. CONCLUSIONS AND DISCUSSION

We have proposed to approximate the Gaussian kernel convolution over the co-occurrence space of Flickr geotags to obtain a location similarity model. This new approach to predict recommendations in a continuous object space can effectively be used to recommend locations matching a user's preference. Recommendations can be made close to the location of the

Fig. 14.   When Ayers Rock in Australia is used as query, the top recommendations in USA West contain many famous rock formations.

user, so that we can suggest landmarks for the next day on a city visit. More interesting, the co-occurrence model can be used to make recommendations in a previously unvisited city or country which is useful while planning a holiday. The bandwidth of the Gaussian kernel controls the size of the target locations, which allows application at a scale of choice (city and country level in this work). The results suggest that recommendations based on the co-occurrence model are both more accurate and more surprising than a ranking based on the prior travel probability. A simple filter to distinguish inhabitants from tourists indicates that touristic behaviour is more informative for the prediction of a user's behaviour in another city.

In this work we have set the weight of all geotags equal, but the proposed model can deal with differently valued data points. We discussed the choice to ignore the number of photos in batch uploads, but a weighting method could be proposed to integrate this information in the amplitude of the data point. Furthermore, the importance of a photo could be estimated on external information sources like the textual tags or the interestingness ranking used by Flickr.

By filtering the set of geotags on the *accuracy* value in the Flickr database we have selected only geotags that are accurate on street level, thereby losing about 40% of the original data. One could argue whether this accuracy filter is necessary if predictions are made on a larger scale (e.g. between-country recommendation). The function that describes a set of geotags is in this work defined as a collection of Dirac delta pulses. To integrate the geotag accuracy into this function, it naturally follows that each geotag could itself be described by a Gaussian

distribution, where the standard deviation is dependent on the accuracy. In this way inaccurate geotags do not influence predictions on small scale, but do contribute on larger scales.

Recommendation evaluation with a training and test set has a drawback. Because of the strongly skewed prior travel distribution most of the locations in a user's test set are well-known popular places. These places will dominate the parameter optimisation of the model, resulting in a personalised model that does not differ much from the prior ranking. The popular locations are however not the most interesting places to recommend, because the user is probably already familiar with them or can easily find them in regular travel guides.

To really evaluate whether a recommender gives interesting, user specific recommendations, manual assessments are inevitable. Using manually annotated locations on both city and country scale we have shown that more strict ranking methods can be used to produce more serendipitous recommendations. A ranking based on cosine similarity can give very interesting and novel recommendations, but also has the possibility of recommending something irrelevant based on data noise. The proposed RankDiff method is more conservative but gives stable good recommendations in all experiments. Based on these results we can assume that these weighting methods will also be more effective in a recommendation system, when the full user profile is used as training data.

APPENDIX A
APPENDIX: FULL 6D KERNEL CONVOLUTION

As indicated in the model description in Section VI, the computation of the co-occurrence model at the prior peak locations is an approximation of the real peaks in the co-occurrence model. To estimate the error introduced by this approximation we have used the mean-shift algorithm to compute the peaks of the full Gaussian kernel convolution on the 6D co-occurrence space for the city pair Berlin-Barcelona at $\sigma = 100\,\mathrm{m}$.

We compare the top-50 similarity relations generated by both methods in the co-occurrence space between Berlin and Barcelona. Using manual evaluation, we find that 44 out of 50 relations uniquely refer to the same landmarks. The median distance of the top-50 peaks in our approximation to the nearest peak in the full convolution is $26\,\mathrm{m}$. The measured peak amplitude at the landmark locations will always be smaller than the nearest peak in the full convolution. We find that the average decay in peak amplitude in the approximation is -2.4%.

The small differences between both models show that the approximation proposed in this work can effectively be used to predict the most co-occurring locations between two cities.

REFERENCES

[1] T. Rattenbury, N. Good, and M. Naaman, "Towards automatic extraction of event and place semantics from flickr tags," in *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*.   New York, NY, USA: ACM, 2007, pp. 103–110. [Online]. Available: http://dx.doi.org/10.1145/1277741.1277762

[2] S. Ahern, M. Naaman, R. Nair, and J. H. Yang, "World explorer: visualizing aggregate data from unstructured text in geo-referenced collections," in *JCDL '07: Proceedings of the 2007 conference on Digital libraries*.   New York, NY, USA: ACM Press, 2007, pp. 1–10. [Online]. Available: http://dx.doi.org/10.1145/1255175.1255177

[3] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: context and content in community-contributed media collections," in *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*.   New York, NY, USA: ACM, 2007, pp. 631–640. [Online]. Available: http://dx.doi.org/10.1145/1291233.1291384

[4] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the world's photos," in *WWW '09: Proceeding of the 18th international conference on World Wide Web*, 2009, pp. 761–770. [Online]. Available: http://www2009.eprints.org/77/

[5] S. S. Lee, D. Won, and D. Mcleod, "Tag-geotag correlation in social networks," in *SSM '08: Proceeding of the 2008 ACM workshop on Search in social media*.   New York, NY, USA: ACM, 2008, pp. 59–66. [Online]. Available: http://dx.doi.org/10.1145/1458583.1458595

[6] P. Serdyukov, V. Murdock, and R. van Zwol, "Placing flickr photos on a map," in *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*.   New York, NY, USA: ACM, 2009, pp. 484–491. [Online]. Available: http://dx.doi.org/10.1145/1571941.1572025

[7] E. Kalogerakis, O. Vesselova, J. Hays, A. A. Efros, and A. Hertzmann, "Image sequence geolocation with human travel priors," in *ICCV '09: Proceedings of the IEEE International Conference on Computer Vision*, 2009.

[8] Y. Takeuchi and M. Sugimoto, "Cityvoyager: An outdoor recommendation system based on user location history," in *Ubiquitous Intelligence and Computing*, J. Ma, H. Jin, L. T. Yang, and J. J. P. Tsai, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, vol. 4159, ch. 64, pp. 625–636. [Online]. Available: http://dx.doi.org/10.1007/11833529_64

[9] T. Horozov, N. Narasimhan, and V. Vasudevan, "Using location for personalized poi recommendations in mobile environments," in *SAINT '06: Proceedings of the International Symposium on Applications on Internet*.   Washington, DC, USA: IEEE Computer Society, 2006, pp. 124–129. [Online]. Available: http://dx.doi.org/10.1109/SAINT.2006.55

[10] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," in *WWW '10: Proceeding of the 19th international conference on World Wide Web*.   New York, NY, USA: ACM, April 2010, p. 10.

[11] Y. Zheng, L. Zhang, X. Xie, and W. Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *WWW '09: Proceedings of the 18th international conference on World wide web*.   New York, NY, USA: ACM, 2009, pp. 791–800. [Online]. Available: http://dx.doi.org/10.1145/1526709.1526816

[12] M. Clements, P. Serdyukov, A. P. de Vries, and M. J. T. Reinders, "Finding wormholes with flickr geotags," in *ECIR 2010, LNCS 5993*, C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. Rüger,

and K. Rijsbergen, Eds. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 658–661. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-12275-0_70

[13] B. Sarwar, G. Karypis, J. Konstan, and J. Reidl, "Item-based collaborative filtering recommendation algorithms," in *WWW '01: Proceedings of the 10th international conference on World Wide Web*. New York, NY, USA: ACM, 2001, pp. 285–295. [Online]. Available: http://dx.doi.org/10.1145/371920.372071

[14] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similarity fusion," in *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY: ACM Press, 2006, pp. 501–508. [Online]. Available: http://jamesthornton.com/cf/paper?paper_id=54

[15] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, January 2003. [Online]. Available: http://dx.doi.org/10.1109/MIC.2003.1167344

[16] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, Aug 1995. [Online]. Available: http://dx.doi.org/10.1109/34.400568

[17] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224–270, 1994.

[18] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511–4515, March 2010. [Online]. Available: http://dx.doi.org/10.1073/pnas.1000488107

[19] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, October 2002. [Online]. Available: http://dx.doi.org/10.1145/582415.582418

[20] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, February 2010. [Online]. Available: http://dx.doi.org/10.1126/science.1177170

[21] O. Celma and P. Herrera, "A new approach to evaluating novel recommendations," in *RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems*. New York, NY, USA: ACM, 2008, pp. 179–186. [Online]. Available: http://dx.doi.org/10.1145/1454008.1454038

[22] P. N. Tan, V. Kumar, and J. Srivastava, "Selecting the right interestingness measure for association patterns," in *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2002, pp. 32–41. [Online]. Available: http://dx.doi.org/10.1145/775047.775053

[23] M. Ohsaki, S. Kitaguchi, H. Yokoi, and T. Yamaguchi, "Investigation of rule interestingness in medical data mining," in *Active Mining*, 2005, pp. 174–189. [Online]. Available: http://dx.doi.org/10.1007/11423270_10